

WHAT IS CLAIMED IS:

1. A method of analyzing a table of data relating to gene expression that varies with time and/or with the changing of environmental conditions, for identifying groups of co-regulated and co-expressed genes, comprising:

defining a clustering criterion of the data of said table;

in function of said clustering criterion, identifying in sub-tables (CLUSTERS) groups of genes that satisfy said clustering criterion;

establishing pair combinations of said sub-tables;

calculating characteristic parameters of the data associated to the groups of genes of each combination;

generating for each combination a characteristic value in function of the characteristic parameters of the groups of genes by a decision algorithm based on soft computing techniques;

identifying the combinations the characteristic value of which is greater than a certain pre-established threshold as "Gene Networks" and discarding combinations of groups of genes the characteristic value of which is smaller than said threshold.

2. The method of claim 1, further comprising the operations of:

defining a number of logic filtering criteria of the data of said table;

for each logic filtering criterion, generating a corresponding filtered sub-table (FILTER) containing data of genes the expression values of which satisfy said logic filtering criterion;
and

establishing pair combinations of sub-tables generated by filtering the data of said table with said filtering criteria and by said clustering.

3. The method of claim 1, wherein said decision algorithm is a fuzzy logic algorithm the antecedents and consequents of which are defined in function of said parameters.

4. The method of claim 1, wherein said parameters are chosen from the group composed of numerical parameters tied to gene expression levels, parameters having a semantic biological meaning, and mixed parameters expressing at the same time a numerical relationship and a semantic meaning.

5. The method of claim 1, wherein said parameters are chosen from a group consisting of:

absolute values of linear correlation coefficients among data associated to pairs of genes;

absolute values of quadratic correlation coefficients among data associated to pairs of genes;

percentage of genes of the combination the final value of gene expression of which is greater than the respective value of initial gene expression;

percentage of genes of the combination the final value of gene expression of which is smaller than the respective value of initial gene expression;

percentage of genes the values of gene expression of which have a same increasing or decreasing time evolution;

percentage of genes that have a maximum value of gene expression in a same condition;
percentage of genes that have ontologies in common;
percentage of genes that have functional domains in common.

6. The method of claim 1, further comprising discarding combinations among sub-tables constituted by a number of genes smaller than a certain pre-established number, considering only once the genes that are comprised in both combined sub-tables.

7. The method of claim 1, wherein said clustering criteria are based on algorithms chosen in the set constituted by agglomerative hierarchic algorithms, non hierarchic Kmeans algorithms, hierarchic sequential Kmeans, non-hierarchic SOM and not exclusive Fuzzy Clustering.

8. The method of claim 5, comprising
calculating correlation coefficients of all pairs of gene sequences of the combination;
subdividing the interval from 0 to 1 in five sub-intervals of equal length and assigning to each of said sub-intervals a respective quantized value of correlation (v_i);
calculating the percentage of correlation coefficients belonging to each sub-interval;
defining for each combination an overall coefficient of linear correlation obtained as arithmetic mean of the quantized values associated to the sub-intervals containing a number of coefficient greater than 50%.

9. The method of claim 5, comprising

calculating coefficients of quadratic correlation of all pairs of gene sequences of a same combination;

defining for each combination a global coefficient of quadratic correlation obtained as arithmetic mean of said correlation values.

10. The method of claim 5, comprising

calculating the percentage of sequence of the combination with a final value of gene expression greater than the initial value of gene expression;

defining a coefficient of global variation of the value of gene expression, comprised between 0 and 1, corresponding to said percentage.

11. The method of claim 5, comprising

calculating the percentage of sequences of the combination with an increasing time evolution;

defining a coefficient relative to the time evolution of the gene expression comprised between 0 and 1 corresponding to said percentage.

12. The method of claim 5, comprising

calculating the percentage of sequences of the combination with a value of gene expression greater than a pre-established threshold in correspondence of a same instant;

defining a coefficient of presence of maximum excursion of the level of gene expression in correspondence of the same instant, comprised between 0 and 1, corresponding to said percentage.

13. An identification system of groups of co-expressed and co-regulated genes, according to the method of claim 1, comprising:

a pre-processing sub-system input with data of a table relative to gene expressions variable with time or relating to different conditions, generating sub-tables (CLUSTERS) of data in groups of genes that satisfy a pre-established clustering criterion;

a processing sub-system of data of said sub-tables (CLUSTERS), generating signals representing characteristic parameters of data associated to genes of a same combination of pairs of said sub-tables;

an intelligent sub-system input with said signals representative of characteristic parameters and outputting data of groups of genes identified as "Gene Networks."

14. The system of claim 13, wherein said intelligent sub-system is a neural fuzzy logic sub-system, trained off-line.